



# 冯裕祺

<https://fengyuqi.xyz>  
853346463@qq.com | +86-182-7839-3372

## 教育

### 东北大学

2020-2023

应用统计, 硕士, 推荐免试

研究方向: 机器学习, 统计学习, 变量选择, 降维方法

CET-4:532, CET-6:489

### 桂林理工大学

2016-2020

GPA: 4.3 / 5.0

应用统计学, 学士

## 简介

有金融风控和金融软件界实习经历  
有丰富的机器学习和统计学习  
具备良好的沟通技巧, 流利的英语能力和良好的组织协调能力

## 专业技能

- Python • R
- Linux • Windows
- MySQL • MongoDB
- Git • Sklearn
- Pytorch • Sklearn • Transformer

## 专业课程

- 高等数理统计 • 概率论
- 时间序列分析 • 机器学习
- 随机过程 • 回归, 多元分析
- 大数据分析

## 兴趣爱好

- 阅读
- 技术研究
- 羽毛球、篮球

## 个人经历

### 第十六届挑战杯全国大学生课外学术科技作品竞赛全国三等奖

2019.11

- 作为项目负责人负责项目的组织, 人员分工和作品撰写, 项目答辩等。

### 第九届挑战杯广西大学生课外学术科技作品竞赛特等奖

2019.6

### 第四届互联网+广西赛区铜奖

2018

### 东北大学优秀研究生干部

2021

- 担任院研究生会组织部部长, 负责策划, 组织学院研究生活动。
- 有丰富的策划组织经验。

### 研究生一等学业奖学金 \* 3

2020 - 2023

## 实习/项目经历

### 中国电子-中电金信研究院 机器学习算法研究员

2022/4 - 2022/8

- 岗位描述: 负责模法师机器人人工智能平台算法调研, 算法模型后端开发, 部署, 监控。
- 调研时间序列预测和聚类算法, 深度学习时间序列预测方法(基于Transformer的Informer模型)进行调研。最终使用Informer, Prophet框架进行时间序列预测模块开发。负责高维时间序列聚类调研, 成功将Informer(长序列预测模型), TICC(高维时间序列聚类算法)开发并集成到模法师机器学习平台。
- 通过将Informer, Prophet算法集成到模法师平台提升了时间序列预测精度, 并且增加了时间序列聚类预测模块, 丰富了平台功能。并结合时间序列预测相关功能撰写相关专利。
- 调研成果:TICC高维时间聚类算法,  
<https://fengyuqi.xyz/2022/06/10/time-series-cluster>  
传统时间序列建模,<https://fengyuqi.xyz/armagarch.html>。

### MINTECH ToB 部门风控实习生

2022/2 - 2022/3

- 公司主页: <https://www.mintechai.com/homeWeb>
- 岗位描述: 负责与巴基斯坦Telenor Microfinance Bank的个人信贷业务风控模型建立以及信贷额度策略调整。
- 工作内容: 从合作方提供的mongoDB数据库中提取用户数据, 输出监控报告, 进行vintage分析和用户画像构建。根据历史用户数据, 清洗数据并构建特征, 针对逾期用户与为逾期用户类不平衡问题使用采样和加权方法进行处理。
- 工作成果: 通过使用过采样方法(SMOTE)和结合自适应加权并结合xgb,logsticregression和随机森林方法进行集成学习, 最终提升模型识别率ks值5%。

### 山西省国资委国有企业绩效评价中心 数据分析员

2021/8 - 2021/9

- 工作内容: 收集、汇总和分析省属国有企业主要经营指标、国有资本金管理和综合绩效评价的基础数据, 研究国有企业绩效评价的体系、方法; 对省属企业年度及任期综合绩效进行分析、研究和评价; 承担省属国企清产核资的技术性工作。
- 模型建立: 根据搜集的山西省省属国有企业数据, 设计相关模型对国有企业进行客观有效的绩效评价, 并初步探索使用机器学习的方法, 对国有企业绩效评价方法进行改革。

## **AN IMPROVED METHOD TO BINARY CLASSIFY THE URTLRA HIGH DIMENSIONAL DATA 第一作者 2021 - 2022**

- 项目描述：对超高维和小样本的分类问题，提出了使用 ball correlation(相比皮尔逊相关系数更好的方法) 结合 abess(最优子集切片算法) 方法的分类方法。
- 方法：通过两阶段方法首先使用 ball 方法对样本进行初步的特征筛选，然后将筛选出的特征使用 abess 方法进行分类。
- 结果：提出的改进模型在 microarray 数据集（DNA 基因序列数据，用于癌症判别诊断）上面的表现经过交叉验证等方法得到的结果明显优于传统的支持向量机，逻辑回归,knn,xgb 和 lightgbm 等方法，分类 AUC 提升 20% 左右，ACC 提升 10%,RECALL 提升 30%,F1 提升 30%，计算时间缩短约 23 分钟。

## **统计变量选择方法汇总 第一作者 2020 - 2021**

- 作品链接：<https://fengyuqi.xyz/book2.pdf>
- 作品描述：汇总整理统计中传统的收缩方法 LASSO(L1 范数惩罚)，ridge(L2 范数惩罚)，SCAD 和 elastic net 等经典收缩方法从原文中总结其原理和特点。并进一步总结出前沿的 screening 方法（针对超高维问题方法）并且将是用统计变量选择方法的应用论文进行汇总。
- 总结：通过对统计变量选择方法按照时间顺序汇总整理，对整个方法发展的历程和思路有了较为清晰的认识。并且通过阅读原文将其原理深度了解，对在实际问题中解决变量选择问题有很好的帮助。

## **呼吸机智能辅助脱机系统建模 项目负责人 2020/11 - 2020/12**

- 项目描述：为了评估 ICU 患者是否撤离呼吸机，基于 MIMIC III 数据库，结合盛京医院搜集到的数据，建立呼吸机撤机预报模型，减少 ICU 中医护工作量，提高撤机准确率。
- 模型建立：利用 Python 语言使用多种算法建立模型，最终利用 Voting 的方法将随机森林、XGBoost 以及多层感知机算法进行集成，在测试集上达到 85% 的准确率。
- 项目成果：该方案已在沈阳市盛京医院开展初步临床对照实验，准确率达到 80%，极大减少了医护的工作量，未来将应用于 ICU 呼吸重症病房。